Economics 102: Analysis of Economic Data
Cameron Spring 2016 April 21
Department of Economics, U.C.-Davis

First Midterm Exam (Version A)

Compulsory. Closed book. Total of 35 points and worth 22.5% of course grade.
Read question carefully so you answer the question.
You are to use only **simple calculations (+, -, /, *, square root) and show all workings**.
Use the calculators provided by the department.
For computations final answers should be to at least four significant digits.
You may remove the formula sheet and the Stata output sheet(s) at end of exam.

Question scores

Question	$1a$	$1b$	$1c$	$1d$	$1e$	$1f$	$2a$	$2b$	$2c$	$2d$	$2e$	$2f$	$3a$	$3b$	$3c$	$3d$	$3e$	*Mult.choice*
Points	1	1	1	1	2	2	1	2	3	1	3	1	3	3	2	2	1	5

1.(a) Suppose there are three observations on variable `x`, respectively, 2, 4 and 5.
What values does variable `y` take given the following Stata command? **Explain your answer.**

```
generate y = x[_n-1]
```

(b) A person invests $1,000 at interest rate 4 per cent per annum. After how many years will the investment be worth $2,000. **Explain your answer.**

(c) Summary statistics show that variable `w` has sample mean 52 and sample variance 4.
Provide the formula for a transformation of variable `w` that has sample mean 0 and sample variance 1.

(d) Average primary support for Donald Trump is 39.1 percent. His support from those with at most high school education is 7.0 percentage points higher than this average. What percentage of those with at most high school education support Donald Trump?

(e) Suppose a person has natural logarithm of earnings equal to 7.15 in 2010 and 7.29 in 2015. What is the approximate percentage change in the level of earnings from 2010 to 2015. **Explain your answer.**

(f) Calculate $\sum_{i=1}^{3}\left(2+\frac{6}{i}\right)$. **Show all workings.**

QUESTION 2 USES STATA OUTPUT GIVEN AT THE END OF THIS EXAM.
In some cases the answer is given directly in the output. In other cases you will need to use the output plus additional computation.
The U.S. data in 2015-16 for graduates with Masters degrees in various disciplines.

2.(a) Does variable `highmeaning` appear to be symmetrically distributed? **Explain your answer.**

(b) If variable `highmeaning` was normally distributed what range of values would you expect 95 percent of observations to lie in? **Explain your answer.**

(c) Provide a **90 percent** confidence interval for the population mean earnings for Masters graduates with 0-5 years work experience. **Show your workings.**

(d) What Stata command would enable you to directly answer part (c)?

(e) The claim is made that population mean earnings for Masters graduates with 0-5 years work experience equals \$60,000. Test this claim at **significance level 0.10**.
State clearly the null and alternative hypothesis and your conclusion.

(f) Suppose we perform a tet of a hypothesis and find that $p = 0.06$. Will we reject or not reject the null hypothesis at significance level 5%? **Explain your answer.**

3.(a) For a sample of four thirty-year olds, the years of completed schooling equalled 12, 15, 13, and 12.
Compute the sample mean and sample variance. **Show all workings.**

(b) Suppose $X = 10$ with probability 0.6, $X = 20$ with probability 0.3 and $X = 30$ with probability 0.1.
Compute the mean, variance and standard deviation of X. **Show all workings.**

(c) Suppose for $X \sim (200, 10^2)$ we form 100 samples of size 25 and obtain 100 sample means $\bar{x}$.
What approximately do you expect the average of the $\bar{x}$ to equal?
What approximately do you expect the standard deviation of the $\bar{x}$ to equal?
Explain your answer.

(d) For the chapter 5 coin toss example provide two approximate histograms: one for x (= 1 if heads and 0 if tails in a single toss), and one for $\bar{x}$ = average of x from 100 coin tosses.

(e) Suppose for the chapter 5 coin toss example we form 400 simple random samples of size 100, from these calculate 400 95% confidence intervals, and find that 350 of these confidence intervals include the true population mean μ. Is this surprising? **Explain your answer.**

Multiple Choice Questions (1 point each)

1. Data on number of doctor visits in 2016 for a sample of 192 individuals is an example of

a. numerical cross-section data

b. numerical panel data

c. categorical cross-section data

d. categorical panel data

e. none of the above

2. The website FRED is a useful source mainly for

a. cross-section data

b. time series data

c. panel data

d. it is not a source for data.

3. An estimator is the best estimator among consistent estimators if

a. it is also unbiased

b. it has the smallest variance

c. it is normally distributed

d. none of the above.

4. Suppose we estimate a parameter θ an obtain an estimate of 4 with standard error 2 based on a random sample of size 100. Then the t test statistic for testing $H_0 : \theta = 2$ against $H_a : \theta \neq 2$ equals

a. 10

b. 2

c. 1

d. none of the above.

5. For the t statistic based on the sample mean $\bar{X}$ to be exactly $T(n-1)$ distributed in small samples it is needs to be assumed that

a. X is $T(n)$ distributed

b. X is $T(n-1)$ distributed

c. X is normally distributed

d. none of the above.

Univariate Data

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad and \qquad s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\bar{x} \pm t_{\alpha/2;n-1} \times (s_x/\sqrt{n}) \qquad and \qquad t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$\text{ttail}(df, t) = \Pr[T > t]$ where $T \sim t(df)$

$t_{\alpha/2}$ such that $\Pr[|T| > t_{\alpha/2}] = \alpha$ is calculated using $\text{invttail}(df, \alpha/2)$.

Bivariate Data

$r_{xy} = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \dfrac{s_{xy}}{s_x \times s_y}$ [Here $s_{xx} = s_x^2$ and $s_{yy} = s_y^2$].

$\widehat{y} = b_1 + b_2 x_i \qquad b_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad b_1 = \bar{y} - b_2\bar{x}$

$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ ResidualSS $= \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ Explained SS = TSS - Residual SS

$R^2 = 1-$ ResidualSS/TSS

$b_2 \pm t_{\alpha/2;n-2} \times s_{b_2}$

$t = \dfrac{b_2 - \beta_{20}}{s_{b_2}} \qquad s_{b_2}^2 = \dfrac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$

$y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$

$\text{E}[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$

Multivariate Data

$\widehat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki}$

$R^2 = 1-$ ResidualSS/TSS $\qquad \bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2)$

$b_j \pm t_{\alpha/2;n-k} \times s_{bj} \qquad and \qquad t = \dfrac{b_j - \beta_{j0}}{s_{b_j}}$

$F = \dfrac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k)$

$F = \dfrac{(\text{ResSS}_r - \text{ResSS}_u)/(k-g)}{\text{ResSS}_u/(n-k)} \sim F(k-g, n-k)$

$\text{Ftail}(df1, df2, f) = \Pr[F > f]$ where F is $F(df1, df2)$ distributed.

F_α such that $\Pr[F > f_\alpha] = \alpha$ is calculated using $\text{invFtail}(df1, df2, \alpha)$.

```
. describe major earlycareer midcareer highmeaning

              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
major           str52   %52s                  Major for the graduate degree
earlycareer     long    %12.0g                Typical earnings with 0-5 years work experience
midcareer       long    %12.0g                Typical earnings 5-10 years work experience
highmeaning     float   %9.0g                 Fraction who say their work makes the world a better
                                                place

. list major earlycareer midcareer highmeaning in 100/101, clean

                      major   earlyc~r   midcar~r   highme~g
100.   Accounting & Business      53500      82300        .44
101.            Epidemiology      54500      82300        .79

. sum major earlycareer midcareer highmeaning

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       major |          0
 earlycareer |        194    56257.22    14535.76      36100     139000
   midcareer |        194    85715.46    22338.64      48100     173000
 highmeaning |        194    .6662887    .1602189        .26        .97

. sum highmeaning, d

        Fraction who say their work makes the world a
                          better place
-------------------------------------------------------------
      Percentiles      Smallest
 1%           .28             .26
 5%           .41             .28
10%           .45             .33       Obs                 194
25%           .54             .34       Sum of Wgt.         194

50%           .68                       Mean           .6662887
                      Largest             Std. Dev.      .1602189
75%            .8             .93
90%           .88             .93       Variance       .0256701
95%           .91             .95       Skewness      -.1739289
99%           .95             .97       Kurtosis       2.164972

. display _n "t_193,.005 = " invttail(193,.005) _n " t_193,.01 = " invttail(193,.01) ///
>  _n "t_193,.025 = " invttail(193,.025) _n  " t_193,.05 = " invttail(193,.05)  ///
>  _n " t_193,.10 = " invttail(35,.10) _n

t_193,.005 = 2.6015425
 t_193,.01 = 2.345824
t_193,.025 = 1.9723317
 t_193,.05 = 1.6527871
 t_193,.10 = 1.3062118
```